RESEARCH TRIANGLE INSTITUTE

STATISTICS RESEARCH DIVISION (STRD)

September 30, 1999

MEMORANDUM

TO: Mike Papp, Shelly Eberly, Jim Flanagan and Ed Rickman

FROM: Mike Riggs and Andy Clayton

SUBJECT: Proposed Protocol for QA/QC of $PM_{2.5}$ Performance Evaluation Program (PEP)

Introduction: Sections I-IV of this memo provide details of a protocol for QA/QC of data for the identification and quantification of bias in the FRM samplers which will be employed in a PEP of the national network of $PM_{2.5}$ samplers. Section V summarizes the results of an evaluation of the proposed methods on PEP data from EPA Regions 5 and 6 in the first two quarters of 1999. Appendix A describes the SAS programs which were written to implement the proposed methodology and Appendix B contains the five SAS programs. All data herein described are from the EPA's PEDS data base.

For the analyses of PEP-FRM samplers to be described, three types of data structures may occur:

A. All the PEP samplers within a region are collocated to a single site for a brief period in January-February of each calendar year.

B. Pairs of PEP samplers are collocated at sites within regions at various times during the remainder of the year

C. Otherwise, single PEP samplers are located at sites within regions, from February-December.

The QA/QC protocols described below evaluate data from structures A and B, only.

I. Regional-Level Evaluation of Data Collected While All PEP Samplers Are Collocated (see pep.QAPP sec. 24.2)

Objectives: To estimate bias of samplers, to determine those samplers that are biased, to estimate repeatability — using data with structure A

For data from the January-February collocation of the entire set of each region's PEP samplers (structure A data), the following ANOVA model will be fit, by region:

(1) $$y_{ij} = \mu + \hat{o}_i + b_j + e_{ij}$$

Where: y_{ij} = the observed $PM_{2.5}$ value of the ith PEP sampler at time j[1]

μ = the fixed $PM_{2.5}$ mean of the "population" of PEP samplers

τ_i = a fixed effect due to the i^{th} PEP sampler

b_j = the random effect of the j^{th} block (sampling time j)

e_{ij} = the random error (assumed normal) of the i^{th} sampler at time j

This model results in an analysis of variance table like the following for each region:

Source of Var.		Mean Square	Comments
Times	MST		Removal of nuisance time effect
Samplers		MSS	Test vs. MSE to determine bias
Residual		MSE	Provides estimate of repeatability if no block x sampler interaction

Following the recommendations of PEP-QAPP, section 24.2, estimates of the i=1 to p $\mu - :_i$ differences (i.e., $\hat{o}_i$), their associated standard errors and T-tests of H_0: the difference=0, may be obtained from the model. Each difference represents the bias in the i^{th} sampler relative to the average of all the other PEP samplers. The T-test will be significant whenever the observed difference is large relative to its standard error. When the significance level (á) is set at 0.05, this is equivalent to requiring that the 95% confidence interval about the difference *not* include zero. If this occurs, the magnitude of the difference will be considered "large" and the associated PEP monitor should be considered out of control.

This recommended approach has the drawback that the linear model estimate of μ is "contaminated" by the effects of any out of control samplers that are included in the regional

---

[1]Throughout the document, it is recommended that the $PM_{2.5}$ concentrations be log-transformed prior to fitting of such models. This recommendation is based on the fact that log-transformed data will tend to have more homogeneous measurement error variability as concentration level changes, while the measurement error variability in the original scale tends to increase with concentration level. Variance homogeneity is assumed in the fitting of such models and in related tests and multiple comparisons procedures.

collocation. To avoid this problem, we propose that pairwise multiple comparisons procedures (e.g., Fisher's LSD, False Discovery Rates and/or Tukey's test) be employed to identify samplers which differ from the majority of the samplers in the region.

Model(1) also provides estimates of the time (b_j) and error (e_{ij}) variance components. Under the assumption that there are no time×sampler interactions, the time×sampler mean square is actually the error mean square. Thus the error variance and the upper bound of the interval estimate of this error variance component are indicative of the repeatability of the PEP samplers. Ideally, these quantities will be sufficiently small. Hence, possible criteria for the maximum acceptable repeatability are:[2]

$$\frac{\sqrt{MSE}}{\hat{\mu}} \# K_1$$

and/or

$$\frac{\sqrt{ucl(MSE)}}{\hat{\mu}} \# K_2$$

Where: MSE = value of the error mean square,
ucl(MSE) = the upper 95% confidence limit on MSE,
$\hat{\mu}$ = model estimate of the mean of the PEP samplers, and K_1 and K_2 are constants.

The lower and upper confidence limits (lcl and ucl) for the repeatability standard deviation are determined as follows:

$$lcl = \sqrt{dfe\,(MSE/\chi^2_{0.975}}$$

and

$$ucl = \sqrt{dfe\,(MSE/\chi^2_{0.025}}$$

Where: dfe = degrees of freedom for the error mean square, and

[2]If log-transformed data are employed, as recommended, then the $\hat{\mu}$ appearing the equations below would be omitted.

$\div^2_p$ = the 100pth percentage point of the chi-square distribution with dfe degrees of freedom.

An alternative model for estimating the repeatability variance of PEP samplers can be used *if the samplers can be assumed to have no bias* (or if those with bias are excluded from the modeling). Within each region, this model is associated with a one-way ANOVA (i.e., a between- versus within-times analysis):

(2) $$y_{hm} = G + P_h + W_{hm}$$

Where: y_{hm} = observed (log-transformed) $PM_{2.5}$ value for sampler m at time h
G = grand mean
P_h = effect of time h
W_{hm} = random (residual) effect of sampler and other components of measurement error for sampler m at time h.

Note that the residuals from model (2), unlike those from model(1), include any sampler biases that may exist. Thus the model fitting for (2) should be performed for data associated with those samplers thought to have negligible bias. The estimate of repeatability (the residual mean square) obtained here represents a pooling of the sampler and error mean squares from model (1). Confidence intervals for the repeatability, based on model (2), can be determined analogously to those for model (1).

II. National-Level Evaluation of Data Collected While All PEP Samplers Are Collocated (see pep.QAPP sec. 24.3)

Objectives: To determine if repeatability is homogeneous across regions or labs — based on data having structure A

Note: The following discussion is described in terms of regions, but "labs" can be substituted for "regions" throughout.

To determine if the variance in the sampler repeated measures is equal among regions, we will first fit the following mixed model:

(3) $$y_{ijk} = \mu + r_k + \hat{o}_{ik} + b_{jk} + e_{ijk}$$

Where: y_{ijk} = the observed (log-transformed) PM_{25} value of the ith PEP sampler at time j in region k

μ = the fixed mean PM_{25} of the "population" of PEP samplers
r_k = a fixed effect due to the k^{th} EPA region
b_{jk} = the fixed effect of the j^{th} sampling time within region k
τ_{ik} = the random effect for sampler i, within region k
e_{ijk} = the random error (assumed normal) of the i^{th} sampler at time j, in the k^{th} region.

(Note: The fitting of this model will produce the same residuals as those from the region-specific model, model (1), of section III.)

To obtain the test of interest, we will use the absolute values of the mixed model residuals (estimates of the e_{ijk}) and fit them using a 2-way ANOVA model involving time-within-region and region:

(4) $$|\hat{e}_{ijk}| = M + R_k + B_{jk} + d_{ijk}$$

Where: M = overall mean
R_k = effect of region k
B_{jk} = effect of time j within region k
d_{ijk} = deviation from model (includes effect of sampler i)

An F test of the region effect (this is Levene's test for variance homogeneity) provides a test of H_0: equal variances among EPA regions. A nonsignificant result will support the conclusion that the repeatability does not vary among regions.

V. Regional-Level Evaluation of Data Collected While Paired PEP Samplers Are Collocated with Routine Network (see pep.QAPP sec. 24.4)

Objectives: To determine if repeatability based on data having structure B is comparable to the repeatability of the structure A data

Estimated variability of collocated pairs of PEP samplers among sites can be determined by modeling the structure B data within a region as a between- versus within-pairs analysis:

(5) $$y_{hm} = G + P_h + W_{hm}$$

Where: y_{hm} = observed (log-transformed) $PM_{2.5}$ value for member m (=1 or 2) of pair h
G = grand mean

P_h = effect of time and location associated with pair h

W_{hm} = residual = effect of sampler and other measurement error on member m of pair h.

Note that the residuals (W_{hm}) will include any sampler biases that may exist. Thus the above model fitting should be performed for data associated with those samplers thought to have negligible bias. The residual mean square provides an estimate of the repeatability variance under these conditions. Confidence intervals based on this mean square can be computed analogously to those for models (1) and (2). A comparison of the residual mean square from model (5) with the residual mean square from model (2)can be made via an F test (two-sided):

$$F = \frac{Model\ 5\ residual\ mean\ square}{Model\ 2\ residual\ mean\ square}$$

This value should be compared to the tabulated F values for dfe(5) and dfe(2) degrees of freedom, where dfe(v) denotes the degrees of freedom associated with the numerator and denominator mean squares, respectively.

IV. National-Level Evaluation of Data Collected While Paired PEP Samplers Are Collocated in the Routine Network (see pep.QAPP sec. 24.5)

Objectives: To determine if repeatability is homogeneous across regions or labs — based on data having structure B.

Note: The following discussion is described in terms of regions, but “labs” can be substituted for “regions” throughout.

To determine if the variation among the paired collocated samplers differs among regions, we used the positive residuals from model (5). Because the samplers are paired, each positive residual is equal to the larger value minus the average of the pair’s values; thus corresponding to each positive residual, there is a negative residual of equal magnitude. Therefore, within a region having H pairs, there are only H independent residuals (one from each pair). We propose to fit a 1-way ANOVA model for a region effect to the positive residuals:

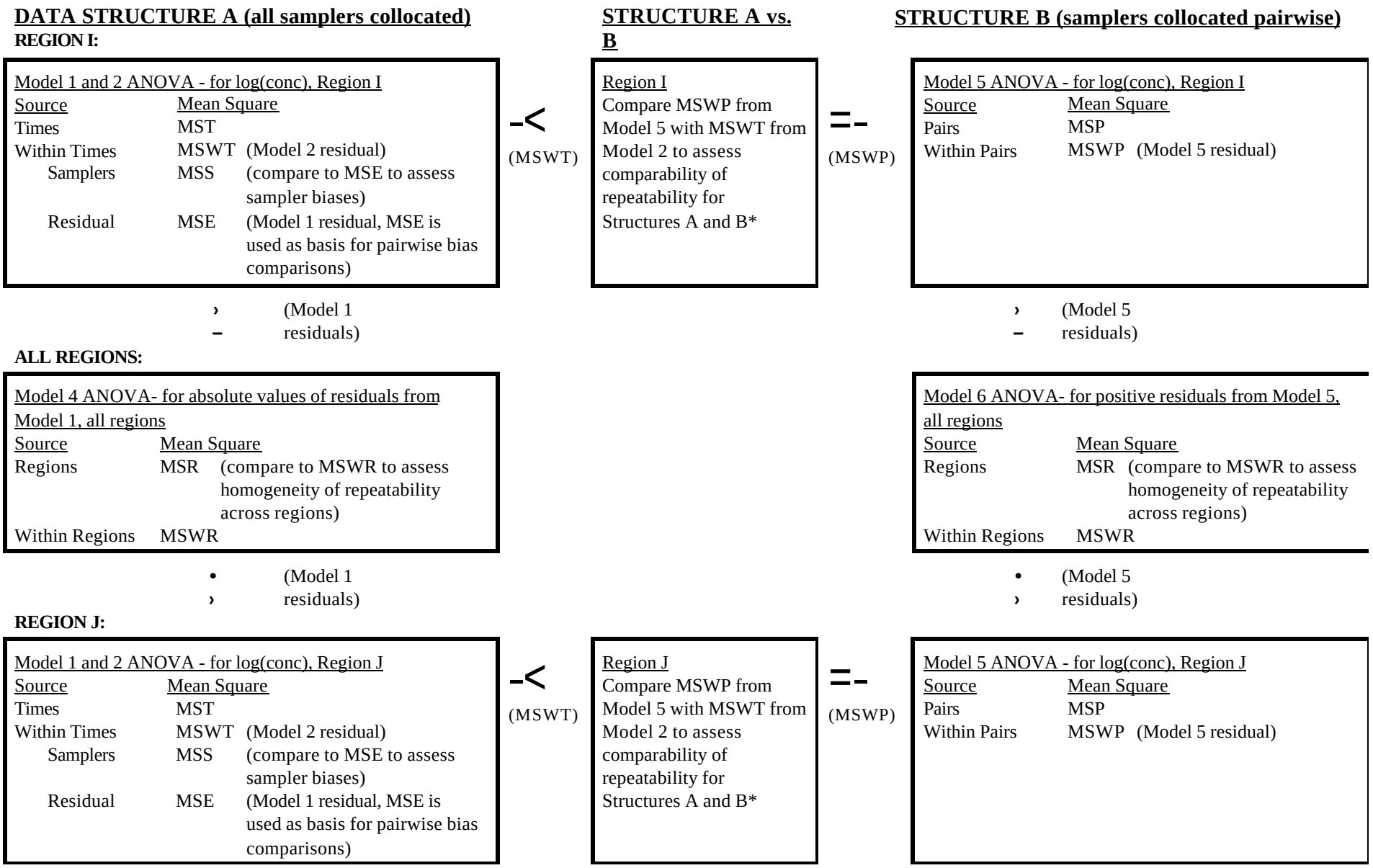
(6) $$\hat{W}_{kh} = ç + ñ_k + ä_{kh}$$

Where: $\hat{W}_{kh}$ = positive residual from model (5) for pair h of region k
ç = overall mean,
$\tilde{n}_k$ = effect of region k, and
$\ddot{a}_{kh}$ = deviation from model.

The F test for the region effect is Levene's test of H_0: equal variances among EPA regions.

It is proposed that the various estimates of repeatability variances for each region — i.e., the residual mean squares from models (1), (2), and (5) — be presented together, along with their corresponding interval estimates. The results would then be further summarized across regions by presenting the results of the Levene's tests, one based on model (4) (associated with the model (1) residuals) and one based on model (6) (associated with the model (5) residuals). Figure 1 provides an overview of this six-step modeling strategy. It illustrates a case with two arbitrary regions (I and J), but the methodology can be extended to any number of regions.

FIGURE 1. OVERVIEW OF ESTIMATION AND TESTING STRATEGY, ILLUSTRATED FOR 2 REGIONS



* The residual mean squares from Models 2 and 5 provide estimates of repeatability variances only if sampler biases do not exist; if differences in MSWP and MSWT are found, it can indicate either a difference in repeatability between Structures A and B or the presence of (more) bias for one of the structures.

V. Results of Preliminary Analysis of Regions 4, 5 & 6 of the PEDS Data, in the First 2 Quarters of 1999.

The data for the winter 1999 omnibus regional collocations in regions 4, 5, and 6 are presented in Table 1. Table 2 summarizes the paired collocations at selected monitoring stations, within regions 5 and 6, during the first two quarters of 1999. Region 4 collocation data (Table 1) were used only in the evaluation of Models 3 and 4. Data from both regions 5 and 6 were used to evaluate all models (1-6). Structure B data (Table 2) were pooled across quarters for assessment of Models 5 and 6. Names of the variables in the tables correspond to PEDS field names: REGION is the EPA region, LAB_ID is the I.D. number of the EPA lab which processed the sample, PE_START is the date on which the field samples were taken and SAMP_ID is
the unique identifier for each $PM_{2.5}$ sampler. The log of the $PM_{2.5}$ mass recovered from the sampler was computed from the PEDS SAMP_MAS variable. Models 1-6 were fit to these data with four SAS macro programs (Appendix A and B).

Assessment of sampler bias via pairwise comparisons among samplers during the winter collocations in Regions 4, 5 and 6 is detailed in Tables 3 - 5. All comparisons are based on the regional error mean-squares estimated from Model 1. The I.D. numbers of the 2 samplers being compared appear in the first column of each table and are followed by the estimated difference (95% confidence limits). The T-statistic is associated with the test of the null hypothesis that the true difference is zero. Simultaneous differences at or near zero, indicate that the samplers in that region were unbiased.. Values in the relative difference column were computed by dividing each observed difference in the untransformed means by the smaller of the paired sampler means and multiplying by 100. Three different methods were used to compute p-values for the pairwise tests. Each employs a different technique to control the Type I error rate. The P-value for Fisher's Least Significant Difference (LSD) is the usual t-test P-value. However, the Type I error is controlled at the significance level of the overall ANOVA test by declaring all individual comparisons to be nonsignifcant (NS) whenever the overall ANOVA F-test is nonsignificant. All three methods employ the pooled MSE and its degrees of freedom in forming the adjusted test statistics.

The Model 1 overall F-test results are provided in the title of each table. Since P>.05 for the overall tests in all three regions, the LSD criterion require us to declare all the pairwise comparisons to be NS, regardless of the values of the individual LSD P-values. Tukey's method utilizes the studentized range distribution to upwardly adjust the individual P-values and assures a family confidence coefficient of at least 1- á; it does not depend on the overall F-test (Neter et al. 1990). Finally, the False Discovery Rate (FDR; Benjamini and Hochberg 1995) controls the false discovery rate but not necessarily the familywise error rate. For the15 pairwise comparisons in Table 3 - 5, all 3 methods lead to the same conclusion; none of the pairwise differences are significantly different from zero. Therefore, the samplers in Regions 4, 5, and 6 can be regarded as unbiased at the time of the winter 1999 regional collocations.

Closer examination of these tables suggests that proper interpretation of the pairwise comparisons will likely require consideration of more than just the adjusted P-values. The

TABLE 1. THE REGIONAL COLLOCATION DATA (DATA STRUCTURE A)

REGION	PE_START	LAB_ID	SAMP_ID	LOG_CONC
4	01/06/99	10	182	3.18221
		10	203	3.19048
		10	204	3.16969
		10	205	3.22287
		10	206	3.21487
4	01/07/99	10	203	3.29213
		10	204	3.32504
		10	206	3.33577
		10	225	3.29584
4	01/11/99	10	203	2.27213
		10	204	2.46810
		10	206	2.21920
5	12/28/98	10	179	3.17805
		10	180	3.16125
		10	194	3.19458
		10	195	3.20680
		10	196	3.22287
		10	200	3.18635
5	12/29/98	10	179	2.48491
		10	180	2.49321
		10	194	2.54160
		10	196	2.54160
		10	200	2.45959
6	01/03/99	4	181	2.17475
		4	183	2.15176
		4	184	2.14007
		4	185	2.21920
		4	186	2.16332
		4	217	2.34181
6	01/06/99	4	181	2.45959
		4	183	2.40695
		4	184	2.36085
		4	185	2.34181
		4	186	2.28238
		4	217	2.39790

TABLE 2. THE PAIRED COLLOCATION DATA (DATA STRUCTURE B)

REGION	LAB_ID	QTR	PE_START	PAIR	SAMP_ID	LOG_CONC
5	10	1	02/23/99	11	180	2.51131
			02/23/99	11	194	2.53852
			03/13/99	10	194	2.23309
			03/13/99	10	195	2.09952
			03/25/99	12	196	2.19088
			03/25/99	12	200	2.18663
5	10	2	04/15/99	14	180	2.64383
			04/15/99	14	195	2.71326
			05/06/99	15	180	2.32975
			05/06/99	15	194	2.44210
			05/09/99	13	179	1.81392
			05/09/99	13	200	1.95060
6	04	1	02/11/99	1	185	2.50079
			02/11/99	1	186	2.48712
			02/17/99	2	185	2.02433
			02/17/99	2	186	2.01357
6	04	2	04/20/99	5	185	2.21882
			04/20/99	5	217	2.25911
			04/27/99	4	185	2.53777
			04/27/99	4	217	2.57976
			05/13/99	3	185	2.08855
			05/13/99	3	186	2.00240
6	04	3	06/05/99	8	185	1.85105
			06/05/99	8	186	1.95616
			06/08/99	6	183	1.90243
			06/08/99	6	184	2.06178
			06/08/99	9	185	2.10390
			06/08/99	9	186	2.24536
			06/10/99	7	183	1.41582
			06/10/99	7	184	1.77586

TABLE 3. MODEL NO. 1 RESULTS: PAIRWISE DIFFERENCES IN LOG(PM2.5 CONC.)
AMONG COLLOCATED REGION 4 PEP SAMPLERS
[OVERALL F-TEST : F(5,4)=0.299, P=0.8911]

SAMPLERS COMPARED	AVG. DIFF.	(95% CONFIDENCE INTERVAL)		T-STAT.	DIFF. IN CONC.	RELATIVE DIFF. IN (%)	LSD P-VALUE	TUKEY'S P-VALUE	FDR P-VALUE
203 - 204	-0.0694	(-0.2572,	0.1185)	-1.03		-7.2	0.3632	0.8883	0.9938
204 - 206	0.0643	(-0.1235,	0.2522)	0.95		6.6	0.3956	0.9130	0.9938
204 - 225	0.0635	(-0.2233,	0.3503)	0.61		6.6	0.5722	0.9837	0.9938
203 - 205	-0.0579	(-0.3447,	0.2289)	-0.56		-6.0	0.6049	0.9889	0.9938
205 - 206	0.0529	(-0.2339,	0.3397)	0.51		5.4	0.6357	0.9926	0.9938
182 - 204	-0.0521	(-0.3389,	0.2347)	-0.50		-5.3	0.6405	0.9930	0.9938
205 - 225	0.0520	(-0.3233,	0.4274)	0.38		5.3	0.7200	0.9980	0.9938
182 - 205	-0.0407	(-0.3660,	0.2847)	-0.35		-4.1	0.7461	0.9988	0.9938
182 - 203	0.0173	(-0.2695,	0.3041)	0.17		1.7	0.8754	1.0000	0.9938
182 - 206	0.0122	(-0.2746,	0.2990)	0.12		1.2	0.9115	1.0000	0.9938
204 - 205	0.0114	(-0.2754,	0.2982)	0.11		1.2	0.9171	1.0000	0.9938
182 - 225	0.0114	(-0.3640,	0.3867)	0.08		1.1	0.9370	1.0000	0.9938
203 - 206	-0.0050	(-0.1929,	0.1828)	-0.07		-0.5	0.9442	1.0000	0.9938
203 - 225	-0.0059	(-0.2927,	0.2809)	-0.06		-0.6	0.9572	1.0000	0.9938
206 - 225	-0.0009	(-0.2877,	0.2859)	-0.01		-0.1	0.9938	1.0000	0.9938

TABLE 4. MODEL NO. 1 RESULTS: PAIRWISE DIFFERENCES IN LOG(PM2.5 CONC.)
AMONG COLLOCATED REGION 5 PEP SAMPLERS
[OVERALL F-TEST : F(5,4)=3.0955, P=0.1481]

SAMPLERS COMPARED	AVG. DIFF.	(95% CONFIDENCE INTERVAL)	T-STAT.	RELATIVE DIFF. IN CONC.(%)	LSD P-VALUE	TUKEY'S P-VALUE	FDR P-VALUE
196 - 200	0.0593	(0.0043, 0.1142)	2.99	6.1	0.0402	0.1965	0.3120
180 - 196	-0.0550	(-0.1100, -0.0000)	-2.78	-5.7	0.0499	0.2368	0.3120
179 - 196	-0.0508	(-0.1057, 0.0042)	-2.56	-5.2	0.0624	0.2857	0.3120
194 - 200	0.0451	(-0.0098, 0.1001)	2.28	4.6	0.0849	0.3663	0.3183
180 - 194	-0.0409	(-0.0958, 0.0141)	-2.06	-4.2	0.1079	0.4404	0.3238
179 - 194	-0.0366	(-0.0916, 0.0184)	-1.85	-3.7	0.1381	0.5261	0.3453
195 - 200	0.0417	(-0.0278, 0.1113)	1.67	4.3	0.1710	0.6067	0.3664
180 - 195	-0.0375	(-0.1070, 0.0321)	-1.50	-3.8	0.2089	0.6855	0.3917
179 - 195	-0.0332	(-0.1027, 0.0363)	-1.33	-3.4	0.2554	0.7643	0.4256
194 - 196	-0.0141	(-0.0691, 0.0408)	-0.71	-1.4	0.5145	0.9698	0.7123
195 - 196	-0.0175	(-0.0871, 0.0520)	-0.70	-1.8	0.5223	0.9721	0.7123
179 - 200	0.0085	(-0.0465, 0.0635)	0.43	0.9	0.6895	0.9966	0.8618
180 - 200	0.0043	(-0.0507, 0.0592)	0.21	0.4	0.8403	0.9999	0.8987
179 - 180	0.0043	(-0.0507, 0.0592)	0.21	0.4	0.8404	0.9999	0.8987
194 - 195	0.0034	(-0.0661, 0.0729)	0.14	0.3	0.8987	1.0000	0.8987

TABLE 5. MODEL NO. 1 RESULTS: PAIRWISE DIFFERENCES IN LOG(PM2.5 CONC.)
AMONG COLLOCATED REGION 6 PEP SAMPLERS
[OVERALL F-TEST : F(5,5)=1.313, P=0.3862]

SAMPLERS COMPARED	AVG. DIFF.	(95% CONFIDENCE INTERVAL)	T-STAT.	RELATIVE DIFF. IN CONC.(%)	LSD P-VALUE	TUKEY'S P-VALUE	FDR P-VALUE
186 - 217	-0.1470	(-0.3108, 0.0168)	-2.31	-15.8	0.0691	0.3335	0.6592
184 - 217	-0.1194	(-0.2832, 0.0444)	-1.87	-12.7	0.1198	0.5015	0.6592
181 - 186	0.0943	(-0.0695, 0.2581)	1.48	9.9	0.1989	0.6896	0.6592
183 - 217	-0.0905	(-0.2543, 0.0733)	-1.42	-9.5	0.2147	0.7190	0.6592
185 - 217	-0.0893	(-0.2531, 0.0744)	-1.40	-9.3	0.2197	0.7278	0.6592
181 - 184	0.0667	(-0.0971, 0.2305)	1.05	6.9	0.3430	0.8835	0.7316
185 - 186	0.0577	(-0.1061, 0.2214)	0.90	5.9	0.4070	0.9296	0.7316
183 - 186	0.0565	(-0.1073, 0.2203)	0.89	5.8	0.4158	0.9346	0.7316
181 - 217	-0.0527	(-0.2165, 0.1111)	-0.83	-5.4	0.4460	0.9496	0.7316
181 - 183	0.0378	(-0.1260, 0.2016)	0.59	3.9	0.5786	0.9869	0.7316
181 - 185	0.0367	(-0.1271, 0.2004)	0.58	3.7	0.5899	0.9886	0.7316
184 - 185	-0.0300	(-0.1938, 0.1337)	-0.47	-3.1	0.6571	0.9953	0.7316
183 - 184	0.0289	(-0.1349, 0.1927)	0.45	2.9	0.6692	0.9961	0.7316
184 - 186	0.0276	(-0.1362, 0.1914)	0.43	2.8	0.6828	0.9968	0.7316
183 - 185	-0.0012	(-0.1649, 0.1626)	-0.02	-0.1	0.9863	1.0000	0.9863

objective is to identify any sampler(s) that is(are) out of control relative to the other collocated samplers. This is facilitated by ordering the pairwise comparisons by the size of the LSD p-values. In Table 5 we see that although none of the adjusted P-values are significant, four of the five largest differences are associated with the same sampler (SAMP_ID=217). Furthermore, the relative differences associated with sampler 217 are substantially larger than those associated with all but one of the other samplers. This may indicate a problem with sampler 217 even though the comparisons are not statistically significant. Similarly, sampler 196 is involved with the three largest differences in Table 4. However, the relative differences associated with sampler 196 are considerably smaller than those associated with sampler 217, so there may be less of a problem with 196. Conversely, a situation may arise wherein only 1 pairwise contrast will be statistically significant. If a sampler is truly out of control it should differ from the majority of the other collocated samplers. A lone significant pairwise contrast will most likely involve the two samplers with the largest and smallest $PM_{2.5}$ concentrations and will be more indicative of a problem with repeatability than of bias.

Table 6 provides a summaries of the analyses of the repeatability of the samplers. Interval estimates of the square-Root-Mean-Square Errors (RMSE) and associated tests of RMSE equality among regions and between data structures A and B are included in the table. If one employs a repeatability criterion that requires the upper confidence limit of the relative standard deviation of the repeatability variance to be # k=0.10, then the upper 95% confidence limits on the Model 1 and 2 RMSE estimates from the log-transformed data must be # 0.10 (refer to pages 2-3). This is the case for Region 5 but not for Region 6. Thus we conclude that repeatability is lower and variability higher in region 6 than we would like. This result casts further doubt on the acceptability of Sampler 217, whose large relative difference values in Table 4 indicate that it is the major contributor to the Region 6 RMSE. Levene's test rejected the null hypothesis for equal variability among the Region 5 and 6 PEP samplers during the winter collocation (Table 6, bottom row). This result is consistent with the pairwise comparisons and the interval estimates of the RMSEs. However, when the test was rerun with Region 4 included (see Table 1), the result was NS ($F_{2,28}$=1.465, P=0.2483; not tabled).

Comparison of the Model 2 estimates of regional variability with the model 5 estimate of the variability among-collocated-pairs, within regions (Table 6, last column), indicates that these 2 components of the error variance are significantly different in region 5 (P=0.0128) but not in region 6 (P=0.1015). The significance of the region 5 test is surprising given the degree of overlap between the two RMSE estimates [0.0291 (0.0200,0.0532) vs. 0.0675 (0.0435,0.1485)]. Finally , Levene's test for equality of the variability of collocated pairs among regions (Model 6; Table 6, last row) was NS, indicating that the variation among paired PEP samplers was about the same from region to region.

References:

ANOVA Models:

Littell, R.C., G.A. Milliken, W.W. Stroup and R.D. Wollfinger. 1996. SAS System for mixed models. SAS Inst. Cary, NC.

Levene's Tests:

Conover, W.J., M.E. Johnson and M.E. Johnson. 1981. A comparative study of tests of homogeneity of variances, with applications to the outer continental shelf bidding data. Technometrics 23(4):351-361.

Snedecor, G.W. and W.G. Cochran. 1980. Statistical Methods, 7th Ed. Iowa State University Press. Ames, Iowa. 507 pp.

Multiple Comparisons:

Benjamini, Y. and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal Stat. Soc. B. 57:289-300.

Neter, J., W. Wasserman and M. H. Kutner. 1990. Applied Linear Statistical Models. 3rd. Ed. Irwin, Homewood, Ill. 1181 pp.

TABLE 6. ASSESSMENT OF SAMPLER REPEATABILITY FOR LOG($PM_{2.5}$ CONCENTRATIONS) BASED ON THE COLLOCATION DATA AND PAIRED DATA OF REGIONS 5 AND 6 IN THE FIRST HALF OF 1999.

REGION	DATA TYPE	MODEL	RESIDUAL DEGREES OF FREEDOM	EST. RESIDUAL STD. DEV. [=RMSE] (95% CONFIDENCE LIMITS)	F-RATIO FOR TEST OF COMPARABLE REPEATABILITY IN DATA TYPES A AND B	
					F*	P-value
5	A A B	1 2 5	4 9 6	0.0198 (0.0119, 0.0569) 0.0291 (0.0200, 0.0532) 0.0675 (0.0435, 0.1485)	5.38	0.0128
6	A A B	1 2 5	2 10 9	0.0637 (0.0398, 0.1563) 0.0685 (0.0479, 0.1202) 0.1047 (0.0720, 0.1911)	2.33	0.1015
Combined	A B	4 6	Levene's test for variance homogeneity over regions: F**= 5.20, p-value= 0.0340 Levene's test for variance homogeneity over regions: F**= 0.14, p-value= 0.9084			

* This F value is given by Model 5 MSE / Model 2 MSE. Its significance is evaluated by comparing it to the F distribution based on degrees of freedom from Models 5 and 2. This is a two-sided test.

**This F value is computed as the (Region MS)/(Residual MS) for the respective model.
Note: Results from Region4 are omitted due to unavailability of Structure B data.

APPENDIX A. Explanation of SAS Programs and Input Data

The data for the regional and paired collocations for the first 2 quarters of 1999 were originally obtained from the EPA PEDS data base as Lotus files. These were converted to a SAS data set. Prior to analysis, this data set was reduced to one containing only the 1999 regional and paired collocation data for Regions 4, 5 and 6 and three new variables were created. These include: the log of the $PM_{2.5}$ concentrations (for each sampler, on each date), a unique identifier for each pair of collocated samples and a data type variable to distinguish data Structure A values (regional collocations) from structure B values (collocated pairs within regions). This data set was the input for the SAS programs used to produce the analyses in Section V. Table 7 summarizes the variables in the complete input data set. A listing of the data was shown in Tables 1 and 2.

Table 7. Variables Contained in the SAS input data set, TESTPEP

Variable	Type	Label
AIRS_SIT	Char	unique ID for sampling location
LAB_ID	Char	LAB ID (EPA region of lab location)
LOG_CONC	Num	Log[PM25]
PAIR	Num	COLLOCATED PAIR I.D.
PE_START	Num	DATE OF FIELD PM25 MEASUREMENT
QTR	Num	QUARTER
DATATYPE	Char	REGIONAL (A) VS. PAIRED COLLOCATION (B) DATA
REGION	Num	EPA REGION OF SAMPLING LOCATION
SAMP_ID	Num	UNIQUE ID FOR SAMPLER

A separate SAS macro program was written for each of the tasks described in sections I-IV. The macros are summarized in Table 8.

Table 8. Summary of the SAS Macro Programs Used to Produce the Analyses in Section 5

Macro Name	Models Evaluated	Call Variables
PEPMAC1.SAS	1 and 2	Input and output data set names and region
PEPMAC2.SAS	3 and 4	Input and output data set names
PEPMAC3.SAS	2 and 5	Input and output data set names and region

PEPMAC4.SAS	5 and 6	Input and output data set names

The macros were called for execution by the SAS program, MACALL.SAS. The macro programs and MACALL.SAS are included in APPENDIX B.

APPENDIX B. SAS Programs